HAL
open science

Is the dinoflagellate Amoebophrya really missing a mtDNA? Letter

Ehsan Kayal, David R Smith

To cite this version:

Ehsan Kayal, David R Smith. Is the dinoflagellate Amoebophrya really missing a mtDNA? Letter. Molecular Biology and Evolution, 2021, 10.1093/molbev/msab041/6132241 . hal-03139036

HAL Id: hal-03139036

https://hal.sorbonne-universite.fr/hal-03139036v1

Submitted on 11 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Is the dinoflagellate *Amoebophrya* really missing a mtDNA?

Letter

Ehsan Kayal[1]*, David R. Smith[2]

[1]Fédération de Recherche 2424 Sorbonne Université & Centre National pour la Recherche Scientifique, Station Biologique de Roscoff, Roscoff, France.

[2]Department of Biology, University of Western Ontario, London, ON, N6A 5B7, Canada.

*Corresponding Author: ehsan.kayal@sb-roscoff.fr

Keywords: Anaerobe, genome loss, oxidative phosphorylation, *Perkinsus*, Syndiniales.

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab041/6132241 by Sorbonne Université user on 11 February 2021

ScholarOne, 375 Greenbrier Drive, Charlottesville, VA, 22901 Support: (434) 964-4100

Abstract

Mitochondrial DNA (mtDNA) is a universal hallmark of aerobic eukaryotes. That is why the recent suggestion by John *et al.* (2019) that the aerobic dinoflagellate *Amoebophrya* sp. strain AT5 (Syndiniales) lacks mtDNA was so remarkable. Here, by reanalysing recently published genomic and transcriptomic data from three *Amoebophrya* strains, we provide evidence of a cryptic, highly reduced mtDNA in this clade. More work is needed before one can definitively say if *Amoebophrya* has or does not have a mtDNA, but for now the data are pointing towards the existence of one. Ultimately, we urge caution when basing supposedly absent genomic features on single line evidences.

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab041/6132241 by Sorbonne Université user on 11 February 2021

ScholarOne, 375 Greenbrier Drive, Charlottesville, VA, 22901 Support: (434) 964-4100

Main text

Mitochondria are, in many ways, inextricably tied to their genomes. Indeed, with one potential exception (discussed below), all aerobic eukaryotes have mitochondrial DNA (mtDNA). But why? Eukaryotic evolution has shown that nearly all of the genes originally housed on the proto-mitochondrial genome have been transferred to the nucleus or lost entirely (Roger et al. 2017). So, why leave behind a handful of mitochondrial genes, all dependent on a complex, metabolically costly gene expression system? Several hypotheses have addressed this question, including the colocation of gene and gene product for redox regulation of gene expression (CoRR) hypothesis (Allen 1993). Moreover, studies of anaerobic eukaryotes with mitochondrial-related organelles (MROs) have shown that it is possible to jettison a mitochondrial genome (Santos et al. 2018), and in extreme cases the mitochondrion itself (Karnkowska et al. 2016). For example, the symbiotic MRO-containing anaerobes *Cryptosporidium* spp. (Apicomplexa), *Henneguya salminicola* (Myxozoa), and *Neocallimastix patriciarum* (Fungi) have all lost their mtDNAs (Rada and Tachezy 2019; Tsaousis and Keithly 2019; Yahalomi et al. 2020), as have the free-living anaerobic protists *Pygsuia biforma* (Obazoa), *Sawyeria marylandensis* (Heterolobosea), and *Mastigamoeba balamuthi* (Amoebozoa) (Gill et al. 2007; Barberà et al. 2010; Stairs et al. 2014). In each of these cases, however, the absence of oxidative phosphorylation has been a key catalyst for forfeiting the mtDNA. What of aerobic eukaryotes?

John *et al.* (2019) (John et al. 2019) recently argued that the syndinialean parasite *Amoebophrya* sp. strain AT5 (Syndiniales, Dinoflagellates) lacks a mitochondrial genome despite having an otherwise conventional aerobic mitochondrion. If true, this would break the paradigm that mitochondrial oxidative phosphorylation is universally linked to the presence of mtDNA (Johnston and Williams 2016). Support for the loss of a mitogenome in *Amoebophrya* strain AT5 was based primarily on cell staining and whole-genome sequencing data. Using confocal laser scanning microscopy in conjunction with SYTOX staining, the authors did not observe DNA in *Amoebophrya* mitochondria but easily identified it in the nucleus. Similarly, their high-coverage total cellular DNA and RNA sequencing of strain AT5 revealed no obvious reads or contigs of mitochondrial origin, with the exception of two small (< 55 nt) fragments of *cox1* (representing the C-terminal moiety of the protein) on scaffolds 46 and 1091, both of which had typical characteristics of nuclear DNA. No other pieces of *cox1* were identified.

ScholarOne, 375 Greenbrier Drive, Charlottesville, VA, 22901 Support: (434) 964-4100

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab041/6132241 by Sorbonne Université user on 11 February 2021

More recently, a genome study of two other *Amoebophrya* strains (A120 and A25) also reported nuclear-encoded *cox1* fragments, corresponding to the metal-binding sites located near the C-terminus of the protein and harbouring mitochondrial transit peptides (Figure 1A) (Farhat et al. 2021). But, here again, the authors were unable to identify any regions encoding the highly conserved oxygen-binding domain of COX1, which is all the more intriguing given that *Amoebophrya* is an aerobic clade (John et al. 2019; Farhat et al. 2021). In the functional annotations of the genomes, Farhat *et al*. did uncover a slew of nuclear-encoded, mitochondrial-targeted proteins involved in mtDNA replication and gene expression, including a homolog of plant organellar DNA polymerases (POP), a DNA-directed RNA polymerase (RNAP), 31 mitochondrial ribosomal proteins, and a monomeric Phenylalanine-tRNA (FARS2) ligase (Figure 1B; Supplementary Table S1), all showing expression during at least one life-cycle stage. This has led to doubts about whether *Amoebophrya* has truly lost its mitochondrial genome (Farhat et al. 2021).

Among the closest known relatives of *Amoebophrya* species for which significant mtDNA sequence data are available are members of the pathogenic dinoflagellate genus *Perkinsus*, whose mtDNAs are highly reduced, fragmented, and undergo induced frameshifts at AGG and CCC positions (Masuda et al. 2010; Zhang et al. 2011). Using *Perkinsus* mitochondrial gene sequences as BLAST queries against the *Amoebophrya* A120 and A25 genome assemblies, we identified in each strain a mtDNA-like contig encoding the missing oxygen-binding domain of COX1 (Figure 1A). These putative mtDNA contigs (Supplementary File S1) had lengths of 3,122 nt (A120) and 4,479 nt (A25), did not map to the nuclear genome scaffolds, and were AT-rich (≥ 67%). This is consistent with the nucleotide composition of other dinoflagellates mitogenomes (Waller and Jackson 2009) but contrasts with the low AT content of the A120 and A25 nuclear genome assemblies (< 53%), including the nuclear DNA-located *cox1* fragments, which are 38-51% AT. We also observed contiguous RNA-seq mapping across the full length of the mitochondrial contigs (max avg. coverage = 295X in A120 and 310X in A25), in line with the pervasive organelle transcription of other clades (Sanitá Lima and Smith 2017). No other mitochondrial-like genes were found on the mtDNA-like contigs, but each contained a terminal inverted repeat, which could be folded into a stem-loop secondary structure (Figure 1D; Supplementary Figure S1). Terminal inverted repeats are a common feature of mitochondrial genomes, including those of Alveolates

ScholarOne, 375 Greenbrier Drive, Charlottesville, VA, 22901 Support: (434) 964-4100

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab041/6132241 by Sorbonne Université user on 11 February 2021

(Waller and Jackson 2009; Smith and Keeling 2013), and are thought to increase the stability of linear DNA molecules (Tomáška et al. 2009).

The portions of the A120 and A25 mtDNA-like contigs that show sequence similarity to *cox1* are relatively small: 277 and 366 nt, respectively, including the final stop codon (Supplementary File S2). Similar to *Perkinsus*, the *cox1* segment of A120 requires a +1 frameshift at a GGC position to keep the coding region intact; this frameshift was not found in the corresponding region of the A25 *cox1* sequence (Supplementary File S3). To better understand how these segments might be functioning within *Amoebophrya* strains A120 and A25, we constructed an artificial COX1 protein sequence (arCOX1) based on the conserved regions of the nuclear and mitochondrial *cox1*-like fragments (Figure 1A and C; Supplementary File S4). The putatively mtDNA-encoded *cox1* pieces from A120 and A25 represent 72 and 45 amino acids, respectively, of arCOX1. The remaining 149-150 amino acids (for A25 and A120, respectively) of arCOX1 come from nuclear-encoded segments. Interestingly, only 39-48% of the nuclear *cox1*-like ORFs (when including the transit peptides) are represented in arCOX1. It is possible that the remaining peptides encoded by the nuclear *cox1*-like fragments are involved in the assembly of the COX1 subunit. Fragmentation and transfer of the C-terminal portion of *cox1* to the nucleus is ancient and widespread in eukaryotes (Gawryluk and Gray 2010), suggesting that mechanisms already exist for the reassembly of a fragmented COX1 into a functional cytochrome *c* oxidase (CIII) complex.

We did re-examine the genome assembly of *Amoebophrya* strain AT5 for putative mtDNA-like contigs, but none were found. Keep in mind, however, that despite the A120 and A25 genome assemblies being of better quality than that of AT5, no mtDNA-like contigs were identified in the genome assemblies. Nevertheless, we did uncover a mtDNA-like sequence from the RNA-seq reads downloaded from the NCBI BioProject PRJNA274490 (Supplementary File S1). This 812 nt sequence has an AT content of 65.4% and encodes a similar portion of COX1 as the mtDNA-like sequences of A120 and A25 (Supplementary Files S2, S3, and S4), providing further support for the existence of mtDNA. Moreover, the *cox1* fragment from AT5 also requires a +1 frameshift at a GGC position similar to that found in A120 but absent in A25 (Supplementary File S2).

Much more work is needed before one can definitively say if members of the *Amoebophrya* genus have or do not have a mtDNA. But, for now, the data are pointing towards the existence of a mitochondrial genome, albeit one that might be among the most reduced of all eukaryotes, even more so than the highly reduced mtDNAs of other myzozoans (Gagat et al. 2017). Further detailed

ScholarOne, 375 Greenbrier Drive, Charlottesville, VA, 22901 Support: (434) 964-4100

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab041/6132241 by Sorbonne Université user on 11 February 2021

cataloguing and analysis of nuclear-encoded, mitochondrial targeted proteins in *Amoebophrya*, especially those involved in mtDNA maintenance and in the electron transport chain (ETC) of the oxidative phosphorylation pathway, will undoubtedly provide even more insights into mitochondrial function within this fascinating clade.

Materials and Methods

For each *Amoebophrya* strain, *cox1* sequences from *Perkinsus marinus* (HQ670240.1) and *P. chesapeaki* (HQ670238.1) were used as BLAST queries (v.2.9.0; evalue 1e-03) against two libraries: one containing unique DNA-seq reads from Farhat *et al.* (Farhat et al. 2021) and another containing meta-transcriptomes assembled for each sampling time using data from Farhat *et al.* (Farhat et al. 2018). Bowtie2 v2.3.5 (--no-discordant --end-to-end) (Langmead and Salzberg 2012) and scripts from the SAMtools v.1.3.1 package (Li et al. 2009) were used to map positive hits to RNA-seq reads from the host (Farhat et al. 2018) in order to filter out contigs with read coverages. The integrity of the remaining contigs were checked by mapping DNA-seq reads and contigs were extended with MITObim v.1.8 (Hahn et al. 2013). The expression of the mtDNA-like contigs over a full infection cycle was monitored with Bowtie2. *cox1* fragment in AT5 was identified by BLAST queries (v.2.9.0; evalue 1e-05) of A120 and A25 sequences against trimmed RNA-seq reads obtained from the NCBI BioProject PRJNA274490. COX1-like fragments were translated using the ExPASy Translate tool from SIB ExPASy Bioinformatics Resources Portal (http://web.expasy.org/translate/) (Artimo et al. 2012) with the Standard Genetic Code and the +1 frameshift in A120 and AT5 were identified manually. Conserved COX1-like blocks for each strain were concatenated to create an artificial COX1 'gene' based on their alignment to reference orthologues using the online version of MAFFT v7 aligner (Katoh et al. 2019). The artificial COX1 genes were realigned to orthologues and the prediction of functional domains were inferred from the human COX1 gene (https://www.uniprot.org/uniprot/P00395). Mitochondrial transfer peptides for nuclear fragments were predicted with TargetP v.2.0 (Armenteros et al. 2019) from the DTU Health Tech online server (https://services.healthtech.dtu.dk/) and MitoProt II v1.101 (Claros and Vincens 1996). Terminal repeats were identified by BLASTn from the NCBI Web BLAST server (https://blast.ncbi.nlm.nih.gov/Blast.cgi) of putative mtDNA contigs against themselves and the secondary structure were predicted using the RNAstructure Web Servers

ScholarOne, 375 Greenbrier Drive, Charlottesville, VA, 22901 Support: (434) 964-4100

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab041/6132241 by Sorbonne Université user on 11 February 2021

(Waller and Jackson 2009; Smith and Keeling 2013), and are thought to increase the stability of linear DNA molecules (Tomáška et al. 2009).

The portions of the A120 and A25 mtDNA-like contigs that show sequence similarity to *cox1* are relatively small: 277 and 366 nt, respectively, including the final stop codon (Supplementary File S2). Similar to *Perkinsus*, the *cox1* segment of A120 requires a +1 frameshift at a GGC position to keep the coding region intact; this frameshift was not found in the corresponding region of the A25 *cox1* sequence (Supplementary File S3). To better understand how these segments might be functioning within *Amoebophrya* strains A120 and A25, we constructed an artificial COX1 protein sequence (arCOX1) based on the conserved regions of the nuclear and mitochondrial *cox1*-like fragments (Figure 1A and C; Supplementary File S4). The putatively mtDNA-encoded *cox1* pieces from A120 and A25 represent 72 and 45 amino acids, respectively, of arCOX1. The remaining 149-150 amino acids (for A25 and A120, respectively) of arCOX1 come from nuclear-encoded segments. Interestingly, only 39-48% of the nuclear *cox1*-like ORFs (when including the transit peptides) are represented in arCOX1. It is possible that the remaining peptides encoded by the nuclear *cox1*-like fragments are involved in the assembly of the COX1 subunit. Fragmentation and transfer of the C-terminal portion of *cox1* to the nucleus is ancient and widespread in eukaryotes (Gawryluk and Gray 2010), suggesting that mechanisms already exist for the reassembly of a fragmented COX1 into a functional cytochrome *c* oxidase (CIII) complex.

We did re-examine the genome assembly of *Amoebophrya* strain AT5 for putative mtDNA-like contigs, but none were found. Keep in mind, however, that despite the A120 and A25 genome assemblies being of better quality than that of AT5, no mtDNA-like contigs were identified in the genome assemblies. Nevertheless, we did uncover a mtDNA-like sequence from the RNA-seq reads downloaded from the NCBI BioProject PRJNA274490 (Supplementary File S1). This 812 nt sequence has an AT content of 65.4% and encodes a similar portion of COX1 as the mtDNA-like sequences of A120 and A25 (Supplementary Files S2, S3, and S4), providing further support for the existence of mtDNA. Moreover, the *cox1* fragment from AT5 also requires a +1 frameshift at a GGC position similar to that found in A120 but absent in A25 (Supplementary File S2).

Much more work is needed before one can definitively say if members of the *Amoebophrya* genus have or do not have a mtDNA. But, for now, the data are pointing towards the existence of a mitochondrial genome, albeit one that might be among the most reduced of all eukaryotes, even more so than the highly reduced mtDNAs of other myzozoans (Gagat et al. 2017). Further detailed

ScholarOne, 375 Greenbrier Drive, Charlottesville, VA, 22901 Support: (434) 964-4100

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab041/6132241 by Sorbonne Université user on 11 February 2021

Claros MG, Vincens P. 1996. Computational Method to Predict Mitochondrially Imported Proteins and their Targeting Sequences. *Eur. J. Biochem.* [Internet] 241:779–786. Available from: http://doi.wiley.com/10.1111/j.1432-1033.1996.00779.x

Farhat S, Florent I, Noel B, Kayal E, Da Silva C, Bigeard E, Alberti A, Labadie K, Corre E, Aury J-M, et al. 2018. Comparative Time-Scale Gene Expression Analysis Highlights the Infection Processes of Two Amoebophrya Strains. *Front. Microbiol.* [Internet] 9:2251. Available from: https://www.frontiersin.org/article/10.3389/fmicb.2018.02251/full

Farhat S, Le P, Kayal E, Noel B, Bigeard E, Corre E, Maumus F, Florent I, Alberti A, Aury J-M, et al. 2021. Rapid protein evolution, organellar reductions, and invasive intronic elements in the marine aerobic parasite dinoflagellate Amoebophrya spp. *BMC Biol.* 19:1.

Gagat P, Mackiewicz D, Mackiewicz P. 2017. Peculiarities within peculiarities ? dinoflagellates and their mitochondrial genomes. *Mitochondrial DNA Part B* [Internet] 2:191–195. Available from: https://www.tandfonline.com/doi/full/10.1080/23802359.2017.1307699

Gawryluk RM, Gray MW. 2010. An ancient fission of mitochondrial cox1. *Mol. Biol. Evol.* 27:7–10.

Gill EE, Diaz-Triviño S, Barberà MJ, Silberman JD, Stechmann A, Gaston D, Tamas I, Roger AJ. 2007. Novel mitochondrion-related organelles in the anaerobic amoeba Mastigamoeba balamuthi. *Mol. Microbiol.* 66:1306–1320.

Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads--a baiting and iterative mapping approach. *Nucleic Acids Res.* [Internet] 41:e129. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23661685

John U, Lu Y, Wohlrab S, Groth M, Janouškovec J, Kohli GS, Mark FC, Bickmeyer U, Farhat S, Felder M, et al. 2019. An aerobic eukaryotic parasite with functional mitochondria that likely lacks a mitochondrial genome. *Sci. Adv.* 5.

Johnston IG, Williams BP. 2016. Evolutionary inference across eukaryotes identifies specific pressures favoring mitochondrial gene retention. *Cell Syst.* [Internet] 2:101–111. Available from: http://dx.doi.org/10.1016/j.cels.2016.01.013

Karnkowska A, Vacek V, Zubáčová Z, Treitli SC, Petrželková R, Eme L, Novák L, Žárský V, Barlow LD, Herman EK, et al. 2016. A eukaryote without a mitochondrial organelle. *Curr. Biol.* 26:1274–1284.

ScholarOne, 375 Greenbrier Drive, Charlottesville, VA, 22901 Support: (434) 964-4100

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab041/6132241 by Sorbonne Université user on 11 February 2021

Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 20:1160–1166.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* [Internet] 9:357–359. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3322381&tool=pmcentrez&rendertype=abstract

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.

Masuda I, Matsuzaki M, Kita K. 2010. Extensive frameshift at all AGG and CCC codons in the mitochondrial cytochrome c oxidase subunit 1 gene of Perkinsus marinus (Alveolata; Dinoflagellata). *Nucleic Acids Res.* 38:6186–6194.

Rada P, Tachezy J. 2019. The Proteome of T. vaginalis Hydrogenosomes. In: Hydrogenosomes and Mitosomes: Mitochondria of Anaerobic Eukaryotes. Springer. p. 177–204.

Roger AJ, Muñoz-Gómez SA, Kamikawa R. 2017. The Origin and Diversification of Mitochondria. *Curr. Biol.* 27:R1177–R1192.

Sanitá Lima M, Smith DR. 2017. Pervasive, Genome-Wide Transcription in the Organelle Genomes of Diverse Plastid-Bearing Protists. *G3.Genes|Genomes|Genetics* [Internet] 7:3789–3796. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28935754%0Ahttp://g3journal.org/lookup/doi/10.1534/g3.117.300290

Santos HJ, Makiuchi T, Nozaki T. 2018. Reinventing an Organelle: The Reduced Mitochondrion in Parasitic Protists. *Trends Parasitol.* [Internet] 34:1038–1055. Available from: https://doi.org/10.1016/j.pt.2018.08.008

Smith DR, Keeling PJ. 2013. Gene conversion shapes linear mitochondrial genome architecture. *Genome Biol. Evol.* [Internet] 5:905–912. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23572386

Stairs CW, Eme L, Brown MW, Mutsaers C, Susko E, Dellaire G, Soanes DM, Van Der Giezen M, Roger AJ. 2014. A SUF Fe-S cluster biogenesis system in the mitochondrion-related organelles of the anaerobic protist Pygsuia. *Curr. Biol.* 24:1176–1186.

Tomáška L, Nosek J, Kramara J, Griffith JD. 2009. Telomeric circles: universal players in telomere maintenance? *Nat. Struct. Mol. Biol.* [Internet] 16:1010–1015. Available from:

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab041/6132241 by Sorbonne Université user on 11 February 2021

ScholarOne, 375 Greenbrier Drive, Charlottesville, VA, 22901 Support: (434) 964-4100

http://www.ncbi.nlm.nih.gov/pubmed/19809492

Tsaousis AD, Keithly JS. 2019. The Mitochondrion-Related Organelles of Cryptosporidium Species. In: Hydrogenosomes and Mitosomes: Mitochondria of Anaerobic Eukaryotes. Springer. p. 243–266.

Waller RF, Jackson CJ. 2009. Dinoflagellate mitochondrial genomes: stretching the rules of molecular biology. *Bioessays* [Internet] 31:237–245. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19204978

Yahalomi D, Atkinson SD, Neuhof M, Sally Chang E, Philippe H, Cartwright P, Bartholomew JL, Huchon D. 2020. A cnidarian parasite of salmon (Myxozoa: Henneguya) lacks a mitochondrial genome. *Proc. Natl. Acad. Sci. U. S. A.* 117:5358–5363.

Zhang H, Campbell DA, Sturm NR, Dungan CF, Lin S. 2011. Spliced leader RNAs, mitochondrial gene frameshifts and multi-protein phylogeny expand support for the genus perkinsus as a unique group of alveolates. *PLoS One* 6:e19933.

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab041/6132241 by Sorbonne Université user on 11 February 2021

ScholarOne, 375 Greenbrier Drive, Charlottesville, VA, 22901 Support: (434) 964-4100

Figure Legends

Figure 1: Putative mitochondrial chromosome in three strains (A120, A25 and AT5) of the parasitoid *Amoebophrya* (Syndiniales, Dinoflagellates). A) Alignment of concatenated COX1-like blocks into artificial COX1 protein sequences (arCOX1) identified in A25, A120 and AT5 to homologues from representative clades (sequence alignment can be found as Supplementary File S4). The nuclear-encoded *cox1*-like fragments (a-f) are derived from predicted proteins for each strain: a = GSA120T00019966001; b = GSA120T00004437001; c = GSA25T00013154001; d = GSA25T00007082001; e = g15932.t1; f = g833.t1. mt: mitochondrial-encoded fragment; nucl: nuclear-encoded fragments. Prediction of metal and oxygen-biding domains are based on the human COX1 gene (uniprot.org/uniprot/P00395). CuB: copper B-binding; Mg/Fe: magnesium and iron-biding; O: oxygen-biding. Colours in alignment are based on the hydropathy scale from NCBI Multiple Sequence Alignment Viewer 1.18.1 (https://www.ncbi.nlm.nih.gov/projects/msaviewer/) where red and blue represent hydrophobic and hydrophilic residues, respectively.

B) Count of tRNA ligase and ribosomal proteins identified in A120 and A25. Numbers in brackets reflect the presence of mitochondrial transit peptide identified by TargetP v.2.0 and/or MitoProt II v1.101.

C) Predicted transcription of COX1-like fragments in the nucleus of A120 and A25. CDS are displayed as blue boxes transcribed in the same direction (left to right) and assembled into contiguous ORFs for nuclear genes, where dark regions correspond to mitochondrial transfer peptides. Orange-shaded regions of nuclear ORFs correspond to conserved blocks of arCOX1 in the alignment (A) and numbers display the GC content.

D) Predicted transcription of COX1-like ORFs from the mtDNA-like contig found in A120 and A25. All genomic features are displayed proportionally to the length of the contigs. DNA-read coverage is displayed for each contig. CDS are displayed as green boxes transcribed in the same direction (left to right). Shaded regions of nuclear ORFs correspond to conserved blocks of arCOX1 in the alignment (A) and numbers display the GC content. Schematic representations of secondary structures of terminal repeats identified in the mtDNA-like contigs (gray boxes at the end of the contigs) were predicted with the RNAstructure Web Servers; for larger images of these structures see Supplementary Files S5-6.

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab041/6132241 by Sorbonne Université user on 11 February 2021

ScholarOne, 375 Greenbrier Drive, Charlottesville, VA, 22901 Support: (434) 964-4100

Supplementary Materials:

Supplementary File S1: Nucleotide sequences of the mtDNA-like contig found in A120, A25 and AT5. AT5 sequence is based on RNA-seq reads downloaded from the NCBI BioProject PRJNA274490.

Supplementary File S2: Nucleotide sequences of the COX1-like segments identified in A120, A25 and AT5. Genes were named according to the nomenclature from Figure 1A. Lower case 'g' in A120 and AT5 sequences corresponds to the presumed +1 frameshift. Sequences for AT5 were identified from the RNA-seq reads downloaded from the NCBI BioProject PRJNA274490. No RNA read was identified corresponding to the predicted AT5_ncox1b_g833.t1 peptide (Supplementary File S3).

Supplementary File S3: Amino acid sequences of the COX1-like segments identified in A120, A25 and AT5. Genes were named according to the nomenclature from Figure 1A and the corresponding protein names from annotated genomes.

Supplementary File S4: Alignment of concatenated COX1-like blocks into artificial COX1 protein sequences (arCOX1) identified in A120, A25 and AT5 to homologues from representative clades.

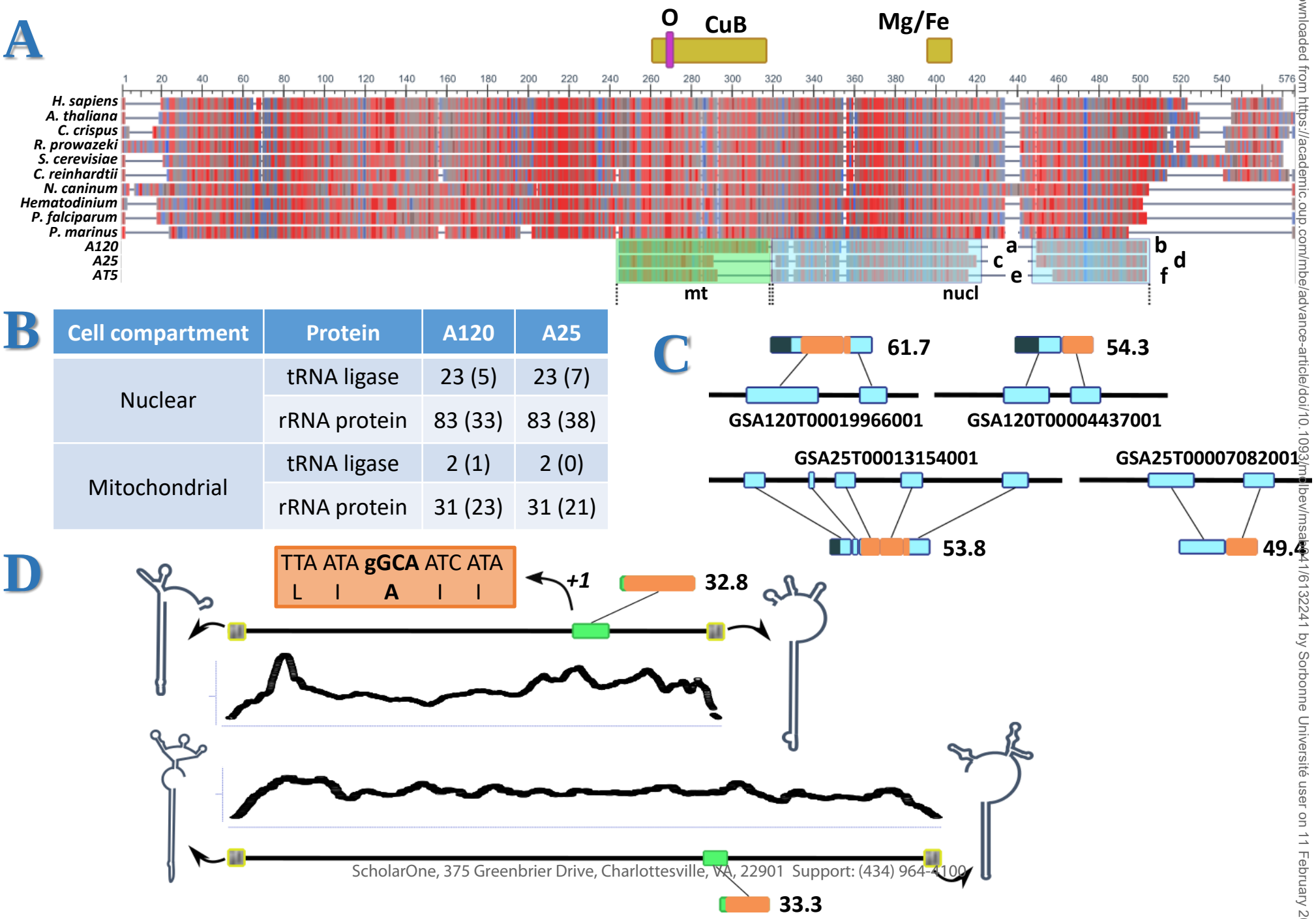
Supplementary File S5: Lowest energy secondary structures of the terminal repeats predicted by the RNAstructure Web Servers on the last 200 nucleotides left (A) and right (B) ends of the mtDNA-like contig of A120.

Supplementary File S6: Lowest energy secondary structures of the terminal repeats predicted by the RNAstructure Web Servers on the last 200 nucleotides left (A) and right (B) ends of the mtDNA-like contig of A25.

Supplementary Table S1: Sequences of selected protein genes involved in the maintenance and expression of a putative mtDNA in the *Amoebophrya* stains A120, A25 and AT5. Sequences for AT5 were identified by BLAST searches (best hit with evalue $1e^{-5}$) of homologues identified in A120 and A25 (Farhat et al. 2021) to the predicted proteome from John *et al.* (2019).

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab041/6132241 by Sorbonne Université user on 11 February 2021

ScholarOne, 375 Greenbrier Drive, Charlottesville, VA, 22901 Support: (434) 964-4100



Cell compartment	Protein	A120	A25
Nuclear	tRNA ligase	23 (5)	23 (7)
	rRNA protein	83 (33)	83 (38)
Mitochondrial	tRNA ligase	2 (1)	2 (0)
	rRNA protein	31 (23)	31 (21)

ScholarOne, 375 Greenbrier Drive, Charlottesville, VA, 22901 Support: (434) 964-4100

Downloaded from https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msab241/6132241 by Sorbonne Université user on 11 February 202